

Exploring metacognitive accuracy in visual search

Joshua S. Redford · Sean Green · Micah Geer ·
Michael Humphrey · Keith W. Thiede

Published online: 6 July 2011
© Psychonomic Society, Inc. 2011

Abstract For decades, researchers have examined visual search. Much of this work has focused on the factors (e.g., movement, set size, luminance, distractor features and proximity) that influence search speed. However, no research has explored whether people are aware of the influence of these factors. For instance, increases in set size will typically slow down target detection; yet no research has measured participants' metacognitive awareness of this phenomenon. The present research explores this area by integrating a visual search task with a metacognitive monitoring paradigm. All of the explored factors influenced search latency. However, all of the factors except target presence influenced ratings. Saliency and suppression are discussed as two possible explanations for the results. Future directions for extending the theory and the practical benefits of this research are also outlined.

Keywords Metacognition · Visual search · Set size · Search asymmetry · Feature overlap

Visual search is a well-studied area that has provided data important to basic and applied research. Basic research on visual search enhances our knowledge regarding the functions of the perceptual system and how our atten-

tional resources are used. Applied research on visual search identifies our limitations in real-world tasks (e.g., tumor detection in X-ray images). Despite the rich visual search literature, research has yet to explore whether people have metacognitive access to their future search abilities. For the sake of brevity, metacognitive accuracy in visual search will be referred to as *metasearch* accuracy. In our view, someone displaying strong metasearch accuracy would be able to accurately identify how different factors (e.g., luminance, movement, distractors) will influence their future search performance.

To measure metasearch accuracy, we developed a visual search paradigm that mirrors traditional methods used for studying metacognitive accuracy. These paradigms measure metacognitive accuracy by requiring participants to try to distinguish well-learned information from poorly learned information. In the typical metacognition paradigm, a participant studies a list of French–English translation equivalents (e.g., *chateau*–*castle*) in preparation for a cued recall test (e.g., *chateau*–?). Prior to the test, the participant rates each item on his or her level of learning—termed the judgment of learning (JOL). The JOL scale provides a range from no learning (e.g., *definitely will not recall*) to complete learning (e.g., *definitely will recall*). The better the participant distinguishes well-learned items (with higher JOLs) from poorly learned items (with lower JOLs), the higher their subsequent metacognitive accuracy. This procedure has been used to assess metacognitive accuracy across a wide range of tasks (for a review of metacognition research, see Dunlosky & Metcalfe, 2009).

Visual search tasks often require participants to view an array of items and indicate whether a target is present. The metasearch paradigm was a hybrid of these two tasks. The metasearch paradigm was composed of two phases—a rating phase and a search phase. In the rating phase, the

J. S. Redford (✉) · M. Humphrey · K. W. Thiede
Boise State University,
1910 University Dr,
Boise, ID 83725, USA
e-mail: JoshRedford@boisestate.edu

S. Green · M. Geer
University at Buffalo, The State University of New York,
Buffalo, NY, USA

computer presented participants with the different search scenes and indicated to the participant whether a target was present. On target trials, the target was presented in the margin of the screen. On no-target trials, the margin showed white space. While studying the scene, the participants rated, on a scale from 1 (*slowest*) to 10 (*fastest*), how quickly they expected to make a decision on the trial when they encountered the scene again during the search phase. So, if the participants expected a scene to be easy, they would provide a fast rating. If they expected a search to be difficult, they would provide a slow rating. After all of the scenes had been rated, participants progressed to the search phase. All of the search scenes were randomly presented. The distractor array remained identical to that viewed during the rating phase. However, the target was placed randomly among the distractor items. We presented the target in the margin and randomly placed it among the distractors during the search phase to avoid the risk that participants would begin to memorize target location, rather than judge the scene's difficulty. Participants searched each scene for the target and responded with a keypress. After the rating and search phases were completed, another cycle (i.e., another rating and search phase) was presented to measure whether any additional improvements in accuracy occurred.

Different cues independently contribute to the accuracy of participants' ratings in studies exploring metacognition. Koriat's (1997) cue utilization framework offers a compelling organization to these cues. Koriat proposed three cue types: intrinsic, extrinsic, and mnemonic. Intrinsic cues refer to characteristics of the to-be-learned items. Participants who provide higher JOLs for closely related word pairs (e.g., *dog–bone*) than for unrelated word pairs (e.g., *dog–book*) are using an intrinsic cue. Extrinsic cues refer to the conditions involved in learning (i.e., factors unrelated to the items). Participants who provide higher JOLs for word pairs presented for 7.5 s than for word pairs presented for 2.5 s are using an extrinsic cue. Mnemonic cues refer to subjective indicators that are used to decide how well-learned something is and its likelihood of future recall. Participants who provide higher JOLs for word pairs that they previously recalled or word pairs that seem more familiar are using a mnemonic cue. Generally, all of these cues are thought to improve metacognitive accuracy, since they provide information regarding which items are more likely to be recalled. However, in certain cases, cues may weaken metacognitive accuracy (see, e.g., Koriat & Bjork, 2005).

Across four experiments, we used the metasearch paradigm to explore whether visual search offered cues to participants as well. In Experiment 1, we manipulated the feature overlap between the target and distractors. If a target shares few features with the distractors in a scene, it becomes easier to detect. Therefore, if participants are

sensitive to this manipulation, they will be faster on search scenes with less feature overlap. In Experiment 2, we used a visual search asymmetry manipulation. Visual search asymmetry refers to search tasks where a member of one group is easy to detect among members of a second ground. However, a member from the second group is difficult to detect among members of the first group. Therefore, if participants are sensitive to this manipulation, they will be faster on search scenes where the easy-to-detect member is the target. In Experiment 3, we manipulated set size. Target detection is easier as the number of distractors is decreased. Therefore, if participants are sensitive to this manipulation, they will be faster on search scenes with fewer distractors. In Experiment 4, the task was designed so that set size was a miscue (i.e., using set size as the cue to predict future search performance would decrease accuracy). The goal of Experiment 4 was to test the robustness of metasearch accuracy with the presence of a miscue. For every experiment, participants were given two cycles of a rating and search phase. In addition to gauging a cue's contribution to metasearch accuracy in novice participants, we included the second cycle to monitor how ratings changed after participants had search experience with the cues.

Experiment 1

In this experiment, we manipulated target–distractor feature overlap. A repeated finding in visual search is that targets are easier to detect if they share few features with the distractor items in a scene. Duncan and Humphreys (1989) demonstrated this phenomenon with a search task that required participants to find an upright *L* among *T*s at different orientations. Specifically, the search was more difficult when the rotated *T*s shared the same features with the *L* (i.e., \perp and \vdash) than when they shared fewer features (i.e., \top and \dashv). We used a similar method for manipulating feature overlap. In Experiment 1, the target was always an upright *T*. In the high-overlap condition, the distractor items were letters with vertical or horizontal lines (Fig. 1a). In the low-overlap condition, the distractor items were letters that did not have vertical or horizontal lines (Fig. 1b). Since feature overlap contributes to the speed of target detection, it may also be a cue that contributes to metasearch accuracy (i.e., participants may use it to determine their future target detection speeds). If feature overlap is an effective cue, we expected participants to predict that they would be faster on the scenes with little feature overlap between the target and the distractors.

Method

Participants Twenty-five students from a psychology course participated in this experiment for extra credit. All participants

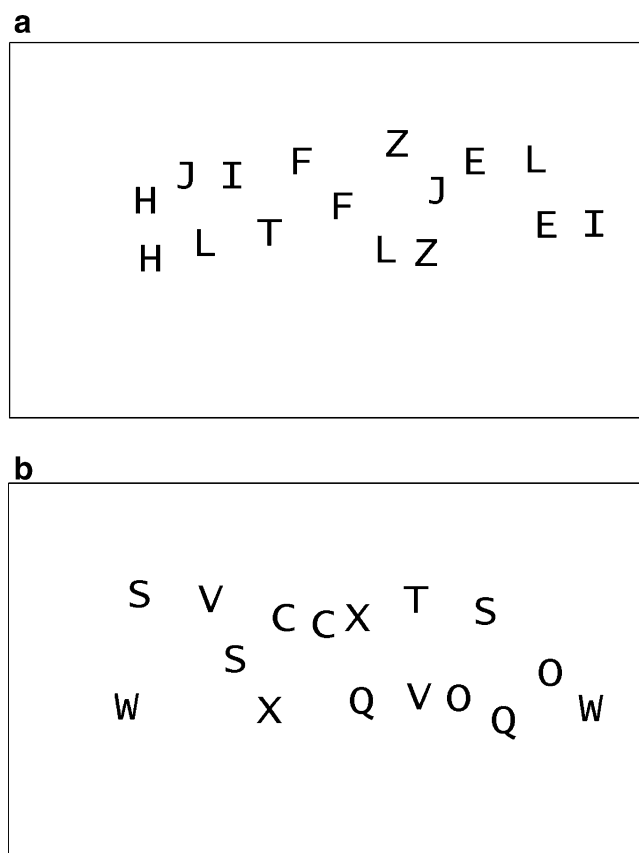


Fig. 1 **a** Screenshot taken from a search scene in the high-overlap condition of [Experiment 1](#). **b** Screenshot taken from a search scene in the low-overlap condition of [Experiment 1](#)

across experiments were treated in accord with APA ethical standards.

Design This experiment used a 2 (cycle: first or second) \times 2 (feature overlap: low or high) \times 2 (target: present or absent) design with all variables manipulated within participants.

Materials Two sets of 24 trials were used. In each set of 24 trials, 12 trials contained a target, and 12 trials did not contain a target. In all cases, the letter *T* served as the target, and each scene consisted of an array of 16 items. Feature overlap was controlled by the features shared between the target and the distractors. Specifically, on high-overlap trials, the distractor letters were based on vertical and horizontal lines (e.g., *H*, *F*, and *E*). On low-overlap trials, the distractor letters displayed no straight vertical or horizontal lines (e.g., *O*, *C*, and *S*). These stimuli were 1.2 cm and were viewed from a distance of about 45 cm—subtending a visual angle of about 1.53°.

Procedure Participants received a verbal description of the task and its requirements. After the verbal description, they received instructions to rate how fast they expected to be,

on a scale from 1 (*slowest*) to 10 (*fastest*). After they finished the ratings, they received instructions to press either “Y” or “N” to respond *yes-target* or *no-target* for the subsequent search trials. After the final search trial, they rated another set of visual search scenes and then searched for the target in the second set of scenes. The purpose of the second cycle was to determine whether metasearch accuracy would improve with practice. After they had finished the second rating–search cycle, participants were debriefed and thanked for their time. A full transcript of the instructions is provided in [Appendix 1](#).

After the participant read the instructions, they searched for a red *T* among black *X*s for ten practice trials. Each trial had a set size of 16 items, with half of the trials containing a target. Participants responded by pressing the “y” key if they thought that a target was present and the “n” key if they thought that no target was present. If correct, the screen flashed green briefly prior to the next trial. If incorrect, the screen flashed red briefly prior to the next trial.

The instructions were identical across experiments and did not suggest how the ratings should be used. Participants then proceeded to the rating phase. [Fig. 2a](#) shows a screenshot of the rating trial. Participants rated 24 trials before moving onto the search phase. Each rating trial displayed the search scene in the top right of the screen, the target (on target trials) in the left margin, and the rating scale across the bottom of the screen. On no-target trials, the left margin was blank. Participants rated each of the 24 trials on a scale ranging from 1 to 10 where 1 indicated the *slowest* search speed and 10 indicated the *fastest* search speed. Participants were allowed as much time as desired on each trial in order to make the most accurate rating possible. To make the data easier to understand and compare with the latency data, we reversed the ratings so that lower values indicated that participants thought that they would be faster. After all of the search scenes were rated, participants moved on to the following set of instruction.

The search scenes were randomly presented. On target trials, the target (originally presented on the left side of the screen during the rating phase) was randomly placed among the distractor items (which were in the same positions as when they were presented during the rating phase). [Figure 2b](#) shows the search trial corresponding to the rating trial shown in [Fig. 2a](#). On no-target trials, the search scene was identical between the rating and search phases. Search scenes were left onscreen until participants used the “y” and “n” keys as they did for the practice trials. Unlike the practice trials, they received no feedback during the search trials. After they provided a *yes* or *no* response, a trial counter displayed their current trial, and then the next search scene was displayed. After they had responded to the

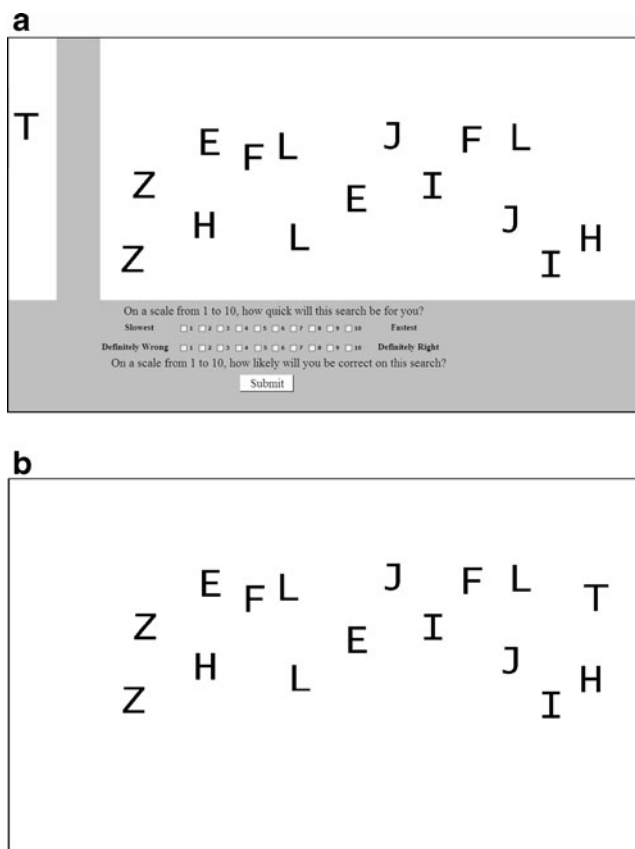


Fig. 2 **a** Screenshot taken from the rating phase of [Experiment 1](#), with the target displayed in the left margin, the search scene displayed to the right, and the rating scale displayed at the bottom of the screen. **b** Screenshot taken from the test phase of [Experiment 1](#). The target is placed randomly among the distractor items

24th search trial, they were given the message, “You’ve now finished a complete set of visual search trials. Please click the button below when you’re ready to start rating the second (and last!) set of visual search trials. Please try to do your best.” The participants then proceeded through another cycle of 24 rating trials and 24 search trials. After they finished the second test phase of 24 trials, they were debriefed and thanked for their time. Most participants took approximately 20 min to complete the entire experiment.

Results and discussion

Metasearch accuracy was based on how sensitive participants were to various factors that could influence search abilities. Therefore, we first present the search analyses to confirm that the manipulations were effective.

Search results For all tests, significance was set at a p value of .05 or less. An ANOVA showed a main effect for cycle, $F(1, 24) = 14.91$, $MSE = 0.45$, for target, $F(1, 24) = 37.90$, $MSE = 0.62$, and for feature overlap, $F(1, 24) = 65.92$,

$MSE = 0.28$. No significant interaction was present across these factors, $F(1, 24) < 1.00$. The absence of a significant three-way interaction was due to a similar influence of cycle across trial types (i.e., search latency gradually decreased across all trial types from the first to the second cycle). Cycle did not interact with feature overlap, $F(1, 24) < 1.27$, $p > .05$, and yielded only a weak, marginally significant interaction with target, $F(1, 24) = 4.23$, $MSE = 0.13$, $p = .05$. In contrast, the interaction between target and feature overlap was robust, $F(1, 24) = 5.40$, $MSE = 0.03$. Follow-up tests confirmed that search was faster for targets trials than for no-target trials and faster for trials with low-feature overlap than for trials with high feature overlap. Follow-up tests also showed that search was faster during the second cycle than during the first cycle. Paired sample t tests were used to further explore the data. Search was faster for target trials than for no-target trials during the first cycle, $t(24) = 6.02$, $SEM = 0.13$, and during the second cycle, $t(24) = 5.11$, $SEM = 0.11$. Search was also faster on target trials with low feature overlap than on trials with high feature overlap, $t(24) = 6.73$, $SEM = 0.07$. Even on the no-target trials, search was faster with low feature overlap than when the distractors shared features, $t(24) = 6.91$, $SEM = 0.11$ (see [Table 1](#), Rows 1 and 2 for latency and rating data). These data are consistent with what we already know about search performance; that is, people find targets faster than they decide that no targets are present in a visual scene, and becomes slower down as feature overlap increases. We now turn to the rating data.

Rating results We reversed the rating data for each experiment to make the ratings easier to compare with the latency data. For example, a maximum rating of 10 was coded as a 1, a 9 was coded as a 2, and so on. An ANOVA showed a marginal effect for cycle, $F(1, 24) = 3.68$, $MSE = 2.75$, $p = .07$, and a main effect for feature overlap, $F(1, 24) = 6.21$, $MSE = 1.84$, but no effect of target on ratings, $F(1, 24) < 1.00$. No significant interaction was present across these factors, $F(1, 24) < 1.00$. Target did not interact with cycle, $F(1, 24) < 1.00$, and yielded only a weak, marginally significant interaction with feature overlap, $F(1, 24) = 3.85$, $MSE = 0.14$, $p = .06$. In contrast, the interaction between cycle and feature overlap was robust, $F(1, 24) = 4.49$, $MSE = 0.19$. Follow-up tests showed no difference in ratings between target and no-target trials. A marginal difference in ratings between the first and second cycles was also present, suggesting that participants generally expected to be faster on the second-cycle trials. Follow-up tests showed a difference in ratings between the low- and high-feature-overlap trials, indicating that participants expected to be faster on the trials with low feature overlap. Paired sample t tests provided

Table 1 Mean target rating and search time by experiment

Experiment	Condition	Target Search (s)	No-Target Search (s)	Target Rating	No-Target Rating
1	Low feature overlap	1.03 (.05)	1.59 (.13) ⁺	3.10 (.35)	3.31 (.42) ⁺
1	High feature overlap	1.52 (.08)	2.32 (.17) ⁺	3.67 (.41)	3.68 (.41) ⁺
2	High salience	1.09 (.07)	2.20 (.13) ⁺	2.31 (.23)	2.69 (.25) ⁺
2	Low salience	1.64 (.08)	1.24 (.06) ⁺	3.33 (.26)	2.47 (.30) ⁺
3	4-item set size	1.52 (.11)	1.97 (.13) ⁺	2.15 (.29)	2.31 (.26) ⁺
3	8-item set size	1.85 (.14)	2.86 (.17) ⁺	3.41 (.27)	3.36 (.29) ⁺
3	16-item set size	2.97 (.14)	4.62 (.27) ⁺	4.87 (.36)	4.44 (.40) ⁺
4	8-item set size	1.44 (.21)	1.68 (.26) ⁺	4.04 (.31)	4.14 (.34) ⁺
4	16-item set size	0.97 (.09)	1.07 (.09) ⁺	2.64 (.30)	2.55 (.32) ⁺

⁺ These data reflect the no-target counterparts to the target trials, (e.g., letters with vertical or horizontal lines)

Standard errors of the means are presented in parentheses

additional information about the interactions. On the basis of their ratings, participants expected to be faster on the trials with low feature overlap than on trials with high feature overlap during the first cycle, $t(24) = 2.22$, $SEM = 0.15$, and the second cycle, $t(24) = 2.56$, $SEM = 0.23$. Participants also predicted that they would be faster on target trials with low feature overlap, as compared with target trials with high feature overlap, $t(24) = 2.60$, $SEM = 0.22$, and on no-target trials with low feature overlap, as compared with no-target trials with high feature overlap, $t(24) = 2.19$, $SEM = 0.17$. The data pertaining specifically to target trials are presented in Fig. 3. The figures present only the target trial data, for two reasons. First, by omitting the no-target data, the figures are easier to understand. Second, and more important, the primary issue of interest throughout this article is the relationship between predictions and search latency. Since ratings

were nearly identical for target and no-target trials, presenting the ratings for no-target trials would be unnecessarily redundant.

Participants made accurate predictions regarding feature overlap. Specifically, they correctly judged that they would be faster when the features present among the distractors were minimally overlapping. They correctly judged that they would be slower when the distractors shared features. This influence of feature overlap on ratings was present during both cycles and whether on target or no-target trials. The discovery that participants made statistically equivalent predictions on target and no-target trials was surprising because of its robust influence on search latency. One possibility is that participants' attentional resources were devoted entirely to the characteristics of the distractor arrays (e.g., distractor configurations, spacing, and so on). Although target presence was indicated by whether the target was presented in the left-side margin, distractor arrays were radically different from trial to trial. A second, related possibility is that participants may have based their ratings strictly on the distractor arrays, since the target–distractor relationship did not require consideration. In other words, feature overlap slowed search time regardless of whether the search scene contained a target. Therefore, since the influence of feature overlap does not require a target–distractor feature comparison, it may further encourage participants to focus on the distractor array alone.

Although the evidence that cues can contribute to meta-search accuracy is promising, the general pattern of disregard toward the target's presence requires additional study. Considering the large influence that target presence has on search latency (i.e., search is faster when the target is present), we expected that target presence would be a default cue used in metasearch ratings. In the next experiment, we used a search asymmetry manipulation to make the target a more salient aspect of the task.

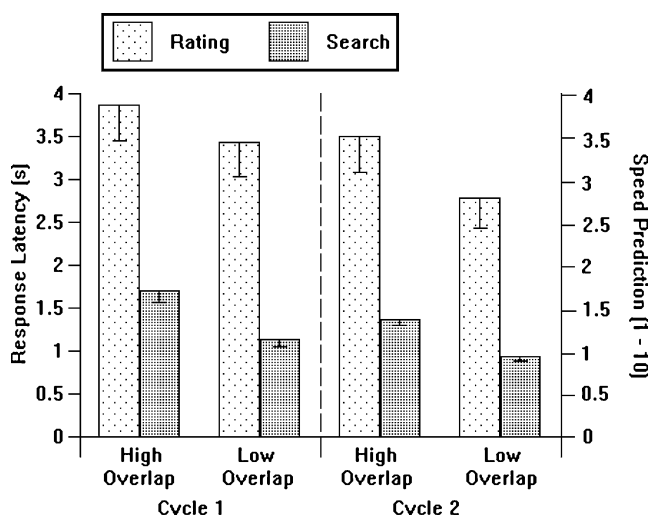


Fig. 3 Summary of the latency and rating data for target trials in Experiment 1

Experiment 2

Experiment 1 used feature overlap in the metasearch paradigm. Search was faster on target trials, and search was faster when the feature overlap was low between the target and the distractors. Participants correctly predicted that they would be faster on the low-feature-overlap trials than on the high-feature-overlap trials. However, they did not consider target presence while making predictions. In the present experiment, we introduce a search asymmetry manipulation. In this task, we used two types of stimuli—a circle (O) and a circle–line symbol (⊙). On high-saliency-target trials, the symbol was the target among the circle distractors. On the low-saliency-target trials, the circle served as the target among the symbol distractors. Target detection will be faster when the trials involved the salient target. We expected the search asymmetry to encourage attention toward the target's presence, for two reasons. First, the search scenes were composed of the same distractor items. The amount of feature overlap among the distractors in the previous experiment served as a useful cue for predicting search speed whether or not the trials involved a target. With uniform distractor arrays, participants are more likely to attend to the relationship between target type and distractor array, since the array, in and of itself, is relatively uninformative. Second, the target proper changes, which should make target presence more noteworthy. To be accurate, participants will be encouraged to view the type of target and, consequently, note whether or not it is a target trial. So, we expected participants to be faster on the target trials relative to the no-target trials, and, within target trials, we expected search to be faster on the salient target trials. We also expected participants to accurately identify both of these cues as being important to search latency. Therefore, we expected participants to predict that they would be faster on the target trials than on the no-target trials and faster on the salient-target trials than on trials where the target was not salient.

Method

Participants Twenty-three students from a psychology course participated in this experiment for extra credit.

Design This experiment used a 2 (cycle: first or second) \times 2 (search asymmetry: target A or target B) \times 2 (target: present or absent) design with all variables manipulated within participants.

Materials Two sets of 24 trials were used. In each set of 24 trials, 12 trials contained a target, and 12 trials did not contain a target. On 6 of the 12 trials, the target was a symbol (⊙) among circles (O). On the other six trials, the

target was the circle (O) among the symbols (⊙). The symbol was the salient target. The circles were 1.3 cm and were viewed from a distance of about 45 cm, subtending a visual angle of about 1.65°. The slightly larger symbols were 1.7 cm and were viewed from a distance of about 45 cm, subtending a visual angle of about 2.16°.

Procedure The procedure was identical to that in [Experiment 1](#).

Results

Search results An ANOVA showed a main effect for cycle, $F(1, 22) = 8.40$, $MSE = 0.16$, for target, $F(1, 22) = 47.14$, $MSE = 0.15$, and for target saliency, $F(1, 22) = 10.73$, $MSE = 0.21$.¹ No significant interaction was present across these factors, $F(1, 22) < 1.00$. As before, the absence of a significant three-way interaction was due to a similar influence of cycle across trial types (i.e., search latency gradually decreased across all trial types from the first to the second cycle). Cycle did not interact with target saliency, $F(1, 22) = 1.56$, $p > .05$, and cycle did not interact with target, $F(1, 22) < 1.00$. A significant interaction occurred between target and target saliency, $F(1, 22) = 124.56$, $MSE = 0.20$. Follow-up tests confirmed that search was faster for targets trials than for no-target trials and faster during the second cycle than during the first cycle. Follow-up tests also showed that search was faster on nonsalient target trials. On the surface, this result is counterintuitive. However, this is comparing the search speed across target and no-target trials. On no-target trials, the absence of a low-salient target takes far longer to determine than the absence of a high-salient target. Paired sample *t* tests were used to further explore the data. Search was faster for target trials than for no-target trials during the first cycle, $t(22) = 6.72$, $SEM = 0.06$, and during the second cycle, $t(22) = 5.94$, $SEM = 0.06$. Search was also faster on high-salient-target trials than on low-salient-target trials, $t(22) = 9.95$, $SEM = 0.05$. As was indicated above, this pattern was reversed for no-target trials. Search was faster on no-target trials when the distractor arrays consisted of (low-salient) circles than when the distractor arrays consisted of (high-salient) symbols, $t(22) = 7.89$, $SEM = 0.12$ (see [Table 1](#), Rows 3 and 4 for latency and rating data). These data again demonstrate that people find targets faster than they decide that no targets are present in a visual scene and also show that search is faster when the target is more salient. We now turn to the rating data, which we expected to align accurately with the search results.

¹ Two search latencies (24.61 and 21.75 s) from a single participant were removed from the data as outliers.

Rating results An ANOVA showed no effect for cycle, $F(1, 22) = 2.63$, $p > .05$, a marginal main effect for target, $F(1, 22) = 3.26$, $MSE = 0.81$, $p = .08$, and a main effect for salience, $F(1, 22) = 28.13$, $MSE = 0.25$. No significant interaction was present across these factors, $F(1, 22) < 1.00$. Target did not interact with cycle, $F(1, 22) < 1.00$, and target did interact with salience, $F(1, 22) = 17.90$, $MSE = 0.99$. Salience also interacted with cycle, $F(1, 22) = 5.95$, $MSE = 0.26$. Follow-up tests showed no difference in ratings between the first and second cycles, a marginal difference between target and no-target trials ($p = .08$), and a significant difference between salient and no-salient trials. Participants predicted that they would be faster on the salient trials. Paired sample t tests provided additional information about the interactions. On the basis of their ratings, participants expected to be faster on the high-salient trials than on the low-salient trials on the first cycle, $t(22) = 5.46$, $SEM = 0.11$, and the second cycle, $t(22) = 2.69$, $SEM = 0.17$. Participants also predicted that they would be faster on target trials with high-salient targets than on those with low-salient targets, $t(22) = 5.61$, $SEM = 0.18$. Understandably, salience did not influence predictions on the no-target trials, $t(22) = 1.54$, $p > .05$. The data pertaining specifically to target trials are presented in Fig. 4.

Experiment 2 continues to demonstrate that participants are capable of identifying useful cues and using them to accurately predict their future search performance. The presence of a target remained an underused cue despite its strong influence on search latency. In contrast, search asymmetry was a cue that contributed heavily to metasearch accuracy. However, metasearch accuracy would have limited utility if it hinges on dramatic changes occurring with stimuli and target type. Most

searches of importance require sensitivity to subtle visual cues (e.g., a slightly darker region on an X-ray indicating a tumor). In the next experiment, the target and distractor items remained constant throughout, with only the number of distractor items (i.e., the set size) changing.

Experiment 3

In Experiment 2, we found evidence that participants were capable of accurately incorporating cues into their search latency ratings. However, the cue in Experiment 2 involved a large amount of perceptual change between target trial types. Both the target and the distractors shifted between a letter (O) and a symbol (Φ). If metasearch accuracy strengthens only when a cue produces pronounced changes in the target and distractors, it offers limited utility. Therefore, we introduced a set size manipulation. By using set size as the cue, we were able to keep the same target for all target trials, and the distractor items did not change in surface appearance. The only factors involved that directly influenced search latency were whether the trial involved a target and the size of the distractor array.

Method

Participants Twenty-two students from a psychology course participated in this experiment for extra credit.

Design This experiment used a 2 (cycle: first or second) \times 3 (set size: 4, 8, or 16 array items) \times 2 (target: present or absent) design with all variables manipulated within participants.

Materials Two sets of 24 trials were used. In each set of 24 trials, 12 trials contained a target, and 12 trials did not contain a target. The target was always an enclosed circle (O) among gapped circles (C). The orientation of the gap was equated at each of four orientations (0°, 90°, 180°, and 270°). The 12 target and 12 no-target trials were partitioned into 4 trials of 4 items, 4 trials of 8 items, and 4 trials of 12 items—meaning that 8 trials consisted of 4 items, 8 trials consisted of 8 items, and 8 trials consisted of 16 items. These stimuli were 0.64 cm and were viewed from a distance of about 45 cm, subtending a visual angle of about 0.5°.

Procedure The procedure was identical to that in Experiment 1.

Results

Search results An ANOVA showed main effects for cycle, $F(1, 21) = 6.78$, $MSE = 1.14$, for target, $F(1, 21) = 52.79$,

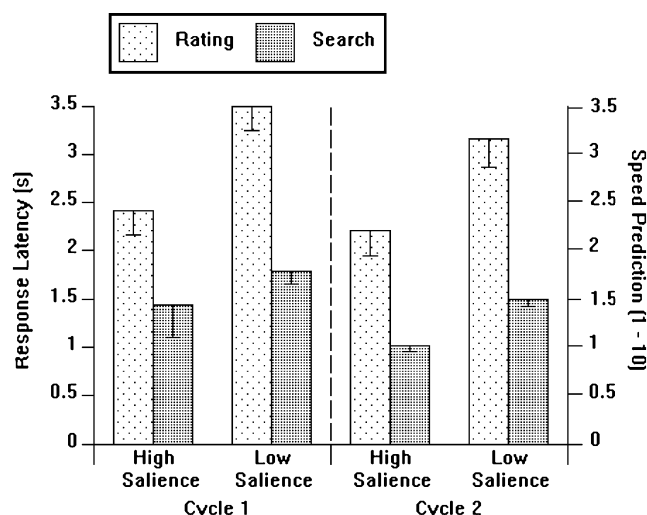


Fig. 4 Summary of the latency and rating data for target trials in Experiment 2

$MSE = 1.33$, and for set size, $F(2, 42) = 198.96$, $MSE = 0.49$. No significant interaction was present across these factors, $F(2, 20) < 1.00$. The patterns observed were the same during both cycles. Cycle did not interact with set size, $F(2, 42) = 2.56$, $p > .05$, and cycle did not interact with target, $F(1, 21) < 1.00$. A significant interaction was present between target and set size, $F(2, 42) = 20.46$, $MSE = 0.38$. Follow-up tests confirmed that search was faster for targets trials than for no-target trials and faster during the second cycle than during the first cycle. Follow-up tests also showed that search was fastest on the smallest set size and slowest on the largest set size. Paired sample t tests confirmed that search was faster for target trials than for no-target trials during the first cycle, $t(21) = 4.79$, $SEM = 0.23$, and during the second cycle, $t(21) = 10.31$, $SEM = 0.09$. An ANOVA confirmed that set size had an effect on target trials, $F(2, 42) = 100.19$, $MSE = 0.13$, and on no-target trials, $F(2, 42) = 129.02$, $MSE = 0.31$. Follow-up tests showed that, for both target and no-target trials, search was fastest for the small set size and slowest for the largest set size (see Table 1, Rows 6–8 for latency and rating data). Unsurprisingly, participants were faster at finding targets and faster at finding targets as the set size decreased. We now consider the rating data to determine whether participants were sensitive to these factors.

Rating results An ANOVA showed no effect for cycle, $F(1, 21) = 2.75$, $p > .05$, or target, $F(1, 21) < 1.00$, but a main effect for set size, $F(2, 42) = 39.63$, $MSE = 3.27$. No significant interaction was present across these factors, $F(2, 42) = 2.10$, $p = .09$. Target did interact with cycle, $F(1, 21) = 7.37$, $MSE = 0.54$, and target did interact with set size, $F(2, 42) = 5.55$, $MSE = 0.34$. Set size also interacted with cycle, $F(2, 42) = 3.41$, $MSE = 0.49$. Follow-up tests showed no difference in ratings between the first and second cycles or between the target and no-target trials. The only significant difference was found for set size. Participants were sensitive to this cue, since they predicted that they would be fastest on the smallest set size and slowest on the largest set size. Paired sample t tests showed no difference in rating between target and no-target trials during the first cycle, $t(21) = 1.95$, $p > .05$, or the second cycle, $t(21) = 1.00$, $p > .05$. An ANOVA confirmed that set size had an effect on target trials, $F(2, 42) = 42.09$, $MSE = 0.96$, and on no-target trials, $F(2, 42) = 29.89$, $MSE = 0.84$. Follow-up tests showed that, for both target and no-target trials, participants predicted that they would be fastest on the small set size trials and slowest on the largest set size trials. The data pertaining specifically to target trials are presented in Fig. 5.

Experiment 1 and Experiment 2 demonstrated that participants are capable of identifying useful cues and using them to accurately predict their future search

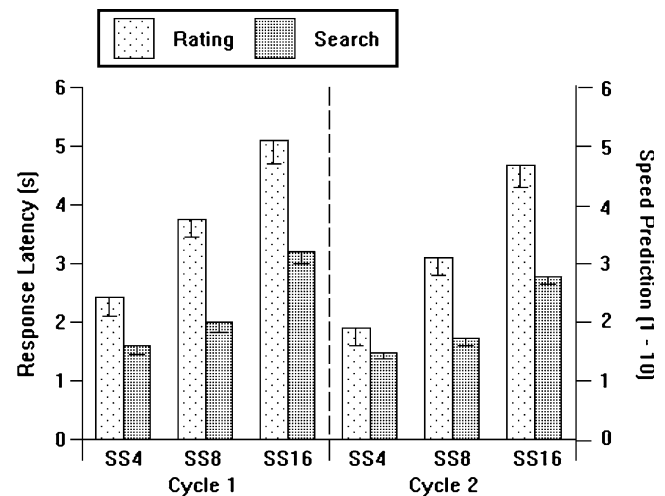


Fig. 5 Summary of the latency and rating data for target trials in Experiment 3

performance. In Experiment 3, the cue was subtler, with the target and distractors remaining similar across trials. Here, the main difference between trial types was the number of distractors in the search array. Participants were still capable of using this cue to accurately predict their future search ability. Across experiments, the presence of a target remained an underused cue—completely neglected in Experiment 1 and Experiment 2 and obtaining only marginal significance in Experiment 2—despite its strong influence on search latency.

These experiments indicate that visual cues that are known to affect search speed and accuracy can contribute to metasearch accuracy. However, they also introduced two other issues. First, can these cues impair metasearch accuracy when they are misleading? In general, we can assume that most individuals have experience searching for items among different set sizes (e.g., a quarter in a change purse) and that a larger set size almost guarantees a more difficult search. Will participants be able to ignore this cue when small arrays actually require more time than large arrays? Second, the previous experiments measured metasearch accuracy but did not identify its source. Put differently, considering that visual search is often a bottom-up process, are participants basing their predictions on concrete cues or on what “feels” easier? To address the first question, we used two cues that contradicted each other. Specifically, we used set size, but the target was more distinct in the large set size condition. If participants were relying on the set size cue, we expected them to display a rating pattern similar to that in Experiment 3. That is, they would predict that they would be faster as the set size decreased even though search would be faster on the large set size trials. If they correctly predicted that they

would be faster on the larger distractor arrays, it would indicate that metasearch accuracy is robust to certain misleading cues. To address the second question, we provided the participants with a questionnaire at the end of the experiment. The purpose of the questionnaire was to encourage the participants to introspect about how they approached both rating and search phases.

Experiment 4

In [Experiment 4](#), we addressed whether metasearch accuracy is robust to miscues or whether metasearch accuracy is inflexibly dependent on cues. In this experiment, we manipulated set size as in [Experiment 3](#), but, this time, set size was a misleading cue. Participants that were relying on set size were likely to make inaccurate predictions. We also provided the participants with a questionnaire to gather information about their introspections during the rating and search phases. A final concern that arose from the previous experiments was that participants may be “experienced” at the first cycle because of the practice session. So, in this final experiment, we removed the practice session and asked the participants to make their ratings without having any prior search experience.

Method

Participants Twenty students from a psychology course participated in this experiment for extra credit.

Design This experiment used a 2 (cycle: first or second) \times 2 (set size: 8 or 12 array items) \times 2 (target: present or absent) design with all variables manipulated within participants.

Materials Two sets of 24 trials were used. In each set of 24 trials, 12 trials contained a target, and 12 trials did not contain a target. The target was always a red-colored *T*. In the 16-item arrays, the distractors were all letter *T*s in other colors, except for red. In the 8-item arrays, the distractors were four black *T*s and three (or four on no-target trials) red *L*s. All of the participants were also given a questionnaire to complete at the end of the experiment (see [Appendix 2](#)). These stimuli were the same size as those described in [Experiment 1](#), so they subtended a visual angle of about 1.53°.

Procedure The procedure was identical to that in [Experiment 1](#), except that the participants completed a questionnaire after they finished the metasearch task.

Results

Search results An ANOVA showed main effects for cycle, $F(1, 19) = 8.72$, $MSE = 0.34$, for target, $F(1, 19) = 17.68$, $MSE = 0.07$, and for set size, $F(1, 19) = 11.69$, $MSE = 1.01$. No significant interaction was present across these factors, $F(1, 19) < 1.00$. Likewise, no significant interactions were present between any of the factors: Cycle did not interact with set size, $F(1, 19) = 1.69$, $p > .05$, cycle did not interact with target, $F(1, 19) = 1.37$, $p > .05$, and target did not interact with set size, $F(1, 19) = 2.32$, $p > .05$. Paired sample *t* tests explored the main effects. Search was marginally faster for target trials than for no-target trials during the first cycle, $t(19) = 1.91$, $p = .07$, $SEM = 0.19$, but not during the second cycle, $t(19) < 1.00$. Search was faster on the larger set size for both target trials, $t(19) = 3.61$, $SEM = 0.13$, and no-target trials, $t(19) = 3.16$, $SEM = 0.19$. The nearly nonexistent influence of target was due to search times being similar for target and no-target trials on the large set size displays since the decision relied on whether the bright, red target was present. In this case, deciding no target is present is as simple as deciding that a target is present. We now consider the rating data to determine whether participants were sensitive to these factors.

Rating results An ANOVA showed no main effect for cycle, $F(1, 19) = 2.16$, $p > .05$, or target, $F(1, 19) < 1.00$, but a main effect for set size, $F(1, 19) = 25.04$, $MSE = 3.58$. The interaction across these factors was significant, $F(1, 19) = 4.41$, $MSE = 0.08$. Target did not interact with cycle, $F(1, 19) = 1.77$, $p > .05$, and target did not interact with set size, $F(1, 19) = 1.71$, $p > .05$, but cycle interacted with set size, $F(1, 19) = 9.12$, $MSE = 0.63$. Follow-up tests showed no difference in ratings between the first and second cycles or between the target and no-target trials. The only significant difference was found for set size. Participants were sensitive to this cue, since they predicted that they would be fastest on the larger set size and slower on the smaller set size. Paired sample *t* tests showed no difference in rating between target and no-target trials during the first cycle, $t(19) < 1.00$, or the second cycle, $t(19) = 1.30$, $p > .05$. Paired sample *t* tests did show that participants predicted that they would be faster on the larger set size for target trials, $t(19) = 4.66$, $SEM = 0.30$, and no-target trials, $t(19) = 5.06$, $SEM = 0.31$ (see [Table 1](#), Rows 9 and 10 for latency and rating data). The data pertaining specifically to target trials are presented in [Fig. 6](#).

The ratings data indicated that metasearch accuracy was a robust phenomenon that is not easily impaired by miscues—even when the cue is something commonly experienced in the real world, such as set size. The presence of metasearch accuracy also indicated that the practice session was unnecessary and, likely, had no influence on the earlier

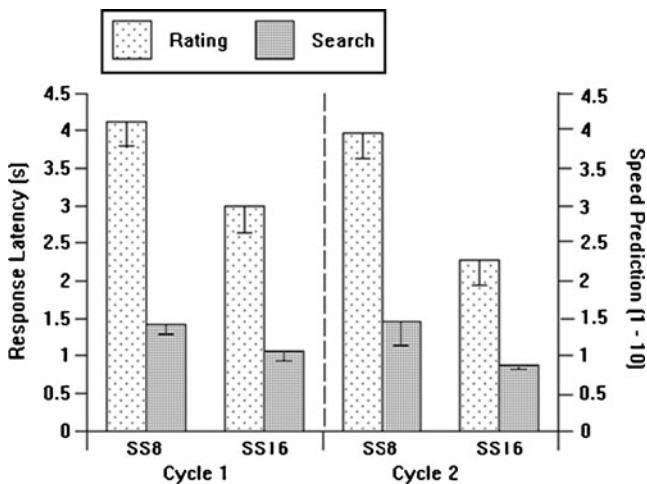


Fig. 6 Summary of the latency and rating data for target trials in Experiment 4

data sets. The questionnaire data provided additional evidence that participants engaged in higher level processing to make their accurate predictions. The majority of participants identified the colors as important to their predictions (i.e., the pastel colors on the 16-item scenes and the other red letters on the 8-item scenes). Other, similarly sophisticated factors that participants considered were whether the trials involved a target and whether the distractors were in close proximity to each other. The participants who did not demonstrate strong metasearch accuracy based their predictions on superficial factors. For the most part, these participants indicated that they guessed on which trials they would be faster. This questionnaire data indicate that strong metasearch accuracy does mean that participants based their ratings on tangible cues present during the rating trials.

General discussion

The metasearch paradigm proved to be useful for exploring monitoring accuracy in visual search. Participants accurately predicted relative search speeds on the basis of feature overlap (Experiment 1), search asymmetry (Experiment 2), set size (Experiment 3), and pop-out (Experiment 4). However, a cue is not guaranteed to contribute to metasearch accuracy even if it influences actual search speeds. Participants routinely failed to consider that no-target trials require more time than did target trials, even though it is arguably the most prevalent finding in the visual search literature.

Prior research offers at least two explanations for why certain cues influence metasearch accuracy and others do not. Metasearch accuracy is not merely an extension of the cue's influence on search time; otherwise, target presence

would have factored into the metasearch ratings. The visual search literature (Nothdurft, 2006) shows that saliency influences search performance. In general, search improves as the target becomes more salient. This trend would suggest that a cue influences speed ratings if it is salient—not necessarily because it does influence speed latency, but because it attracts the participants' attention. If one finds their attention drawn to some part or characteristic of a search scene or target, it is reasonable to assume that this particular salient dimension will influence search latency; hence, it evokes higher ratings. The cue's saliency would explain why there was a marginal effect for target in Experiment 2. The cue depended heavily on target type and the target's relationship to the distractors—thereby, making it more salient than in the other experiments. Metasearch accuracy may have been present in each of the experiments, since feature overlap (Experiment 1), search asymmetry (Experiment 2), and set size manipulations (Experiment 3) produced obvious perceptual changes that were easy for participants to monitor. This explanation can also account for the data pattern in Experiment 4. Despite the same set size change manipulation, participants disregarded this cue and overwhelmingly based their ratings on whether the search scene would involve bright, red distractors instead of pastel-colored distractors. In short, the contrast in distractor colors was more salient than the set size manipulations. The only area where this explanation has difficulty is accounting for the repeated demonstration that target presence had no influence on ratings. Whether a target was present on the rating trial is a salient cue and should have influenced ratings. A related idea that can account for all the data, including the general absence of a target's influence on ratings, stems from recent work by Bjork and his colleagues (Anderson, Bjork, & Bjork, 1994; Storm, Bjork, Bjork, & Nestojko, 2006).

Bjork and his colleagues (Anderson et al., 1994; Storm et al., 2006) found that practice recalling information suppresses other, related information. For instance, participants who study category–exemplar pairs (*fruit–banana*) and then practice recalling half of the studied items will do better on the practiced items. However, the interesting discovery is that this practice suppresses the participants' ability to recall the nonpracticed items, in addition to facilitating recall of the practiced items. A similar process may be happening during the rating phase of the metasearch paradigm. Saliency is still important to ratings, but the saliency also suppresses other, less salient aspects of the search scene. As before, metasearch accuracy was present in each of the experiments because the manipulations (e.g., search asymmetry) were salient changes. Set size had no influence in Experiment 4 because its influence was suppressed by the more salient changes in distractor colors. Most important, target presence never significantly influ-

enced ratings, because, in every experiment, other changes were more salient and, consequently, suppressed the important role that target presence has in search latency. This suppression explanation requires additional testing in future research, since many questions remain. At this point, it is unclear whether suppression is actually occurring. If it is, we also need to understand whether there is a way to avoid or eliminate suppression when multiple cues are relevant. In addition to this, there are several other areas pertaining to metasearch that require investigation.

Future directions

Koriat's cue utilization framework (1997) offered a helpful system for organizing cues in learning tasks. By organizing these cues on the basis of their source, subsequent research was able to study them in isolation. A similar system of classification would help future research in which the cues that contribute to metasearch accuracy are explored. Intrinsic cues (i.e., characteristics of the items) may include factors such as set size. Extrinsic cues (i.e., conditions involved in the task) may include factors such as display times. Mnemonic cues (i.e., subjective indicators of performance) may include previous efforts to search for the target. However, this area poses additional challenges not found in traditional metacognition paradigms. For example, the definition of these cues depends on whether one views the "item" as the target or the search scene. If focused on the target and target characteristics, instead of the scene, certain cues such as set size are more accurately defined as an extrinsic cue. It is premature to impose a framework on these cues, but it may require both a scene-level framework and a target-level framework.

In addition to identifying a system that facilitates the study of cues, research should explore the stability of metasearch accuracy. [Experiment 4](#) suggests that participants are able to ignore miscues in a search task. However, a miscue may be more detrimental when the appropriate cue is less salient. For example, perhaps set size may impair metasearch accuracy when the relevant cue is target orientation. Also, will participants continue to display metasearch accuracy when they are rating scenes under conditions of high cognitive load? Previous research has shown that identifying category members (Waldron & Ashby, 2001), visual search (Emrich, Al-Aidroos, Pratt, & Ferber, 2010), and making metacognitive judgments (Griffin, Wiley, & Thiede, 2008) are all impaired under conditions of high cognitive load. Therefore, one could expect cognitive load to have similar effects on metasearch accuracy as well.

Equally important is how well metasearch findings generalize to real-world scenarios involving visual search. Accurate visual search is vital in a number of areas. Baggage

screeners have to identify potential threat objects in baggage images. Medical doctors have to identify potential tumors in X-ray images. Failure to find these "targets" can produce grave consequences. If these individuals are able to identify cues that hinder search, they may be better able to make appropriate adjustments (e.g., spending more time searching, revisiting a previous search scene, seeking additional help, and so on). It may also be the case that strong searchers are those individuals who are especially attuned to the important cues that contribute to strong search, just as strong readers are those individuals who are especially attuned to the important cues that contribute to strong reading (Thiede, Griffin, Wiley, & Anderson, 2010).

As was stated previously, this article reflects an initial step into an area that has potential for both basic and applied research. We have established that different cues contribute in varying degrees to metasearch accuracy and that predictions do involve higher level considerations (e.g., scene configurations). Future research needs to work toward refining the theory that best accounts for these data and extending the paradigm to real-world scenarios.

Acknowledgements Many thanks to John Dunlosky and Jeremy Wolfe for their guidance and support of this research.

Appendix 1

Introductory instructions

Welcome to the visual search task. This task will involve searching for a specific object—called the target—among several other objects. Try to be as quick as possible while maintaining your accuracy. Speed is as important as accuracy in this task.

Before you start the task, you will be shown a series of split screens. On the left side will be the target and on the right will be the objects that you should ignore. For these split-screen trials, we would like you to estimate how quick you expect to be and how likely you will be correct.

Speed will be on a scale from 1 to 10 (1 means you will be very slow to detect the target, 10 means you will be very fast to detect the target).

Likelihood of being correct will be on a scale from 1 to 10 (1 means you are certain to be wrong, 10 means you are certain to be correct).²

When you are ready, click the button below to move onto the next screen.

² Although we collected accuracy data, participants were near perfect in their detection rates. Consequently, only the latency data is discussed.

Practice instructions

To understand the visual search task, you will now be given a few practice trials.

On each practice trial, you will be searching for a T that may or may not be among the X's.

Press the "y" key if you think the target is present (target-Yes). Press the "n" key if you think the target is not present (target-No). If correct, the screen will briefly turn green. If incorrect, the screen will briefly turn red. This feedback is intended to give you an understanding of the task, but will not be present for the real search task.

Press the enter key to start the visual search task.

Rating instructions

Congratulations, you've completed the practice trials. You will be rating your upcoming trials in two ways. Each display will show the search scene on the right side of the screen. The target, if present, will be displayed on the left side of the screen.

You will rate how quick you think you will be to detect the target when it is among the nontargets in the visual scene from 1 (very slow) to 10 (very fast).

Second, you will rate how likely you think you will be correct in deciding if the target is present in the visual scene from 1 (certain to be incorrect) to 10 (certain to be correct).

[In the first three experiments, we also asked participants to make accuracy predictions. However, on virtually every trial, every participant searched until he or she made the correct decision. Therefore, we focus the analyses and discussion on the more informative speed ratings.]

Important: If the left side of the screen is blank, that particular trial will be a "no target" trial. You should then make your ratings based on how quick you will make a "no target" response (and the likelihood that you will correctly say "no target" on this trial). After the ratings, the search trials will show the nontarget items in the same configurations, but the targets (on target trials) will be placed randomly somewhere among the nontarget items.

Press the button below to start rating the visual search trials.

Search instructions

Congratulations! You have finished these ratings. Now you will have the opportunity to search for the targets on these same trials. Targets, if present on the search trial, will be somewhere among the distracter items. Press the "y" key if you think the target is present. Press the "n" key if you

think the target is absent. Work as quickly as you can while maintaining your accuracy. Good luck!

To get ready, move the cursor to the edge of the screen and place your fingers on the "y" and "n" key. Press enter to start searching.

Appendix 2

1. What did you use to guide your ratings during the first phase?
2. Why did you predict you would be faster in these scenes?
3. What did you use to guide your ratings during the second phase?
4. Why did you predict you would be faster in these scenes?
5. Did your strategies change at all between your rating or search phases?

References

- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1063–1087.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks: Sage.
- Emrich, S. M., Al-Aidroos, N., Pratt, J., & Ferber, S. (2010). Finding memory in search: the effect of visual working memory load on visual search. *Quarterly Journal of Experimental Psychology*, 63, 1457–1466.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self explanation: concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, 36, 93–103.
- Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370.
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 187–194.
- Nothdurft, H. (2006). Saliency and target selection in visual search. *Visual Cognition*, 14, 514–542.
- Storm, C. S., Bjork, E. L., Bjork, R. A., & Nestojko, J. F. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? *Psychonomic Bulletin & Review*, 13, 1023–1027.
- Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, 47, 331–362.
- Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, 8, 168–176.